PROTEASE VARIANTS

Field of the Invention

The present invention relates to a novel protease 3D structure, as well as variants of a parent protease, in particular variants of amended properties, such as improved thermostability and/or amended temperature activity profile. The invention also relates to DNA sequences encoding such variants, their production in a recombinant host cell, as well as methods of using the variants, in particular within the field of animal feed and detergents. The invention furthermore relates to methods of generating and preparing protease variants of amended properties. Preferred parent proteases are Nocardiopsis proteases, such as proteases comprising the mature peptide parts of SEQ ID NOs: 2, 4, 6, 8, 10, and 21.

Background of the Invention

Protease sequences derived from strains of Nocardiopsis are disclosed in WO 88/03947, WO 01/58276, and DK 1996 00013 ("Protease 10," SEQ ID NOs: 1-2).

JP 2003284571-A discloses, as SEQ ID NOs: 2 and 1, the amino acid sequence and the corresponding DNA sequence, respectively, of a protease derived from Nocardiopsis sp. TOA-1 (FERM P-18676). The sequences have been entered in the GENESEQ database as GENESEQP no. ADF43564, and GENESEQN no. ADF43563, respectively.

JP 2-255081-A discloses a protease derived from Nocardiopsis sp. strain OPC-210 (FERM P-10508), however without sequence information. The strain is no longer available, as the deposit was withdrawn.

DD 200432|8 discloses a proteolytic preparation derived from Nocardiopsis dassonvillei strain ZIMET 43647, however without sequence information. The strain appears to be no longer available.

Additional Nocardiopsis protease sequences are disclosed in PCT/DK04/000433 ("Protease 08," SEQ ID NOs: 9-10 herein); PCT/DK04/000434 ("Protease 11," SEQ ID NOs: 5-6 herein); PCT/DK04/000432 ("Protease 18," SEQ ID NOs: 3-4 herein); and PCT/DK04/000435 ("Protease 35," SEQ ID NOs: 7-8 herein).

It is an object of the present invention to provide alternative proteases, in particular for use in animal feed and/or detergents, in particular novel and improved protease variants, preferably of amended properties, such as improved thermostability and/or a higher or lower optimum temperature.

Summary of the Invention

The present invention relates to a variant of a parent protease, comprising a substitution in at least one position of at least one region selected from the group of regions

consisting of: 6-18; 22-28; 32-39; 42-58; 62-63; 66-76; 78-100; 103-106; 111-114; 118-131; 134-136; 139-141; 144-151; 155-156; 160-176; 179-181; and 184-188; wherein

(a) the variant has protease activity; and

(b) each position corresponds to a position of amino acids 1 to 188 of SEQ ID NO: 2; and

(c) the variant has a percentage of identity to amino acids 1 to 188 of SEQ ID NO: 2 of at least 60%.

The present invention also relates to isolated nucleic acid sequences encoding the protease variant and to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing and using the protease variants.

Brief Description of the Figures

Figure 1 is a multiple alignment of Protease 10, Protease 18, Protease 11, Protease 35 and Protease 08 (the mature peptide parts of SEQ ID NOs: 2, 4, 6, 8 and 10, respectively), also including a protease variant of the invention, viz. Protease 22 (amino acids 1-188 of SEQ ID NO: 21); and

Figure 2 provides the coordinates of the novel 3D structure of Protease 10 (amino acids 1 to 188 of SEQ ID NO: 2) derived from Nocardiopsis sp. NRRL 18262.

Detailed Description of the Invention

Three-dimensional Structure of Protease 10

The structure of Protease 10 was solved in accordance with the principles for X-ray crystallographic methods as given, for example, in X-Ray Structure Determination, Stout, G.K. and Jensen, L.H., John Wiley & Sons, Inc. NY, 1989. The structural coordinates for the crystal structure at 2.2 Å resolution using the isomorphous replacement method are given in Fig. 2 in standard PDB format (Protein Data Bank, Brookhaven National Laboratory, Brookhaven, CT). The PDB file of Fig. 2 relates to the mature peptide part of Protease 10 corresponding to residues 1-188 of SEQ ID NO: 2.

Molecular Dynamics (MD)

Molecular Dynamics (MD) simulations are indicative of the mobility of the amino acids in a protein structure (see McCammon, JA and Harvey, SC., (1987), "Dynamics of proteins and nucleic acids", Cambridge University Press). Such protein dynamics are often compared to the crystallographic B-factors (see Stout, GH and Jensen, LH, (1989), "X-ray structure determination", Wiley). By running the MD simulation at, e.g., different temperatures, the temperature related mobility of residues is simulated. Regions having the highest mobility or flexibility (here isotropic fluctuations) may be suggested for random mutagenesis. It is here understood that the high mobility found in certain areas of the protein, may be thermally

improved by substituting these residues.

Using the programs CHARMM (Accelrys) and NAMD (University of Illinois at Urbana-Champaign) the Protease 10 structure described above was subjected to MD at 300 and 400K. Starting from the coordinates of Figure 2 hydrogen and missing heavy atoms were built using CHARMM procedures HBUILD and IC BUILD respectively. Then the structure was minimized using CHARMM Conjugate Gradients (CONJ) minimization procedure for a total of 200 steps. The protein was then put on a 70 X 70 X 70 Angstrom box and solvated with TIP3 water molecules. A total of 11124 water molecules were added and then minimized, keeping the protein coordinates fixed, using CHARMM Adopted Basis Newton Raphson (ABNR) minimization procedure for 20000 steps. The system was then heated to the desired temperature at a rate of 1K every 100 steps using the NAMD software. After an equilibration of 50 picoseconds, an NVE ensemble MD was run for 1 nanosecond, both steps done with the software NAMD. A cut-off of 12 Angstrom was used for the non-bonded interactions. Periodic boundary conditions were used after the solvation step and for all the subsequent ones. The isotropic root mean square (RMS) fluctuations were calculated with the CHARMM procedure COOR DYNA.

The following suggested regions for mutagenesis result from MD simulations: From residue 160 to 170, from residue 78 to 90, from residue 43 to 50, from residue 66 to 75, and from residue 22 to 28.

Strategy for Preparing Variants

Regions of amino acid residues, as well as individual amino acid substitutions, were suggested for mutagenesis based on the 3D-structure of Fig. 2 and the alignment of the five known proteases (upper five rows of Fig. 1), mainly with a view to improving thermostability.

The following regions were suggested, cf. claim 1: 6-18; 22-28; 32-39; 42-58; 62-63; 66-76; 78-100; 103-106; 111-114; 118-131; 134-136; 139-141; 144-151; 155-156; 160-176; 179-181; and 184-188.

At least one of the following positions of the above regions are preferably subjected to mutagenesis, cf. claim 3; 6; 7; 8; 9; 10; 12; 13; 16; 17; 18; 22; 23; 24; 25; 26; 27; 28; 32; 33; 37; 38; 39; 42; 43; 44; 45; 46; 47; 48; 49; 50; 51; 52; 53; 54; 55; 56; 58; 62; 63; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 78; 79; 80; 81; 82; 83; 84; 85; 86; 87; 88; 89; 90; 91; 92; 93; 94; 95; 96; 97; 98; 99; 100; 103; 105; 106; 111; 113; 114; 118; 120; 122; 124; 125; 127; 129; 130; 131; 134; 135; 136; 139; 140; 141; 144; 145; 146; 147; 148; 149; 150; 151; 155; 156; 160; 161; 162; 163; 164; 165; 166; 167; 168; 169; 170; 171; 172; 173; 174; 175; 176; 179; 180; 181; 184; 185; 186; 187; and/or 188.

Contemplated specific variants are listed in the claims, viz. variants of Protease 10, Protease 18, Protease 11, Protease 35 as well as Protease 08 in claims 4 and 15; variants of

Protease 10 in claim 16; variants of Protease 18 in claim 17; variants of Protease 11 in claim 18; variants of Protease 35 in claim 19; and variants of Protease 08 in claim 20.

The various concepts underlying the invention are also reflected in the claims as follows: Stabilization by disulfide-bridges in claims 5 and 6; proline-stabilization in claims 7-8; substitution of exposed neutral residues with negatively charged residues in claims 9-10; substitution of exposed neutral residues with positively charged residues in claims 11-12; substitution of small residues with bulkier residues inside the protein in claim 13; and regions proposed for mutagenesis following MD simulations in claim 14.

The term "at least one" means "one or more," viz., e.g. in the context of regions: One, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, or seventeen; or, in the context of positions or substitutions: One, two, three, four, five, and so on, up to e.g. ninety.

In a particular embodiment, the number of regions proposed for and/or subjected to mutagenesis is at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, or at least seventeen.

In another particular embodiment, the number of regions proposed for and/or subjected to mutagenesis is no more than one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, or no more than seventeen.

Polypeptides Having Protease Activity

Polypeptides having protease activity, or proteases, are sometimes also designated peptidases, proteinases, peptide hydrolases, or proteolytic enzymes. Proteases may be of the exo-type that hydrolyse peptides starting at either end thereof, or of the endo-type that act internally in polypeptide chains (endopeptidases). Endopeptidases show activity on N- and C-terminally blocked peptide substrates that are relevant for the specificity of the protease in question.

The term "protease" is defined herein as an enzyme that hydrolyses peptide bonds. This definition of protease also applies to the protease-part of the terms "parent protease" and "protease variant," as used herein. The term "protease" includes any enzyme belonging to the EC 3.4 enzyme group (including each of the thirteen subclasses thereof). The EC number refers to Enzyme Nomenclature 1992 from NC-IUBMB, Academic Press, San Diego, California, including supplements 1-5 published in Eur. J. Bio-chem. 1994, 223, 1-5; Eur. J. Biochem. 1995, 232, 1-6; Eur. J. Biochem. 1996, 237, 1-5; Eur. J. Biochem. 1997, 250, 1-6; and Eur. J. Biochem. 1999, 264, 610-650; respectively. The nomenclature is regularly supplemented and updated; see e.g. the World Wide Web (WWW) at http://www.chem.qmw.ac.uk/iubmb/enzyme/index.html.

Proteases are classified on the basis of their catalytic mechanism into the following

groups: Serine proteases (S), Cysteine proteases (C), Aspartic proteases (A), Metallo proteases (M), and Unknown, or as yet unclassified, proteases (U), see Handbook of Proteolytic Enzymes, A.J.Barrett, N.D.Rawlings, J.F.Woessner (eds), Academic Press (1998), in particular the general introduction part.

In particular embodiments, the parent proteases and/or the protease variants of the invention and for use according to the invention are selected from the group consisting of:

(a) Proteases belonging to the EC 3.4.-.- enzyme group;

(b) Serine proteases belonging to the S group of the above Handbook;

(c1) Serine proteases of peptidase family S2A; and

(c2) Serine proteases of peptidase family S1E as described in Biochem.J. 290:205-218 (1993) and in MEROPS protease database, release 6.20, March 24, 2003, (www.merops.ac.uk). The database is described in Rawlings, N.D., O'Brien, E. A. & Barrett, A.J. (2002) MEROPS: the protease database. Nucleic Acids Res. 30, 343-346.

For determining whether a given protease is a Serine protease, and a family S2A protease, reference is made to the above Handbook and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

Protease activity can be measured using any assay, in which a substrate is employed, that includes peptide bonds relevant for the specificity of the protease in question. Assay-pH and assay-temperature are likewise to be adapted to the protease in question. Examples of assay-pH-values are pH 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12. Examples of assay-temperatures are 30, 35, 37, 40, 45, 50, 55, 60, 65, 70, 80, 90, or 95°C. Examples of protease substrates are casein, such as Azurine-Crosslinked Casein (AZCL-casein). Examples of suitable protease assays are described in the experimental part.

Parent Protease

The parent protease is a protease from which the protease variant is, or can be, derived. For the present purposes, any protease can be used as the parent protease, as long as the resulting protease variant is homologous to Protease 10, i.e. the protease derived from Nocardiopsis sp. NRRL 18262 and comprising amino acids 1-188 of SEQ ID NO: 2.

In a particular embodiment the parent protease is also homologous to Protease 10.

In the present context, homologous means having an identity of at least 60% to SEQ ID NO: 2, viz. amino acids 1-188 of the mature peptide part of Protease 10. Homology is determined as generally described below in the section entitled Amino Acid Homology.

The parent protease may be a wild-type or naturally occurring polypeptide, or an allelic variant thereof, or a fragment thereof that has protease acticity, in particular a mature part thereof. It may also be a variant thereof and/or a genetically engineered or synthetic

polypeptide.

In a particular embodiment the wild-type parent protease is i) a bacterial protease; ii) a protease of the phylum Actinobacteria; iii) of the class Actinobacteria; iv) of the order Actinomycetales v) of the family Nocardiopsaceae; vi) of the genus Nocardiopsis; and/or a protease derived from vii) Nocardiopsis species, such as Nocardiopsis alba, Nocardiopsis antarctica, Nocardiopsis composta, Nocardiopsis dassonvillei, Nocardiopsis exhalans, Nocardiopsis halophila, Nocardiopsis halotolerans, Nocardiopsis kunsanensis, Nocardiopsis listeri, Nocardiopsis lucentensis, Nocardiopsis metallicus, Nocardiopsis prasina, Nocardiopsis sp., Nocardiopsis synnemataformans, Nocardiopsis trehalosi, Nocardiopsis tropica, Nocardiopsis umidischolae, or Nocardiopsis xinjiangensis.

Examples of such strains are: Nocardiopsis alba DSM 15647 (wild-type producer of Protease 08), Nocardiopsis dassonvillei NRRL 18133 (wild-type producer of Protease M58-1 described in WO 88/03947), Nocardiopsis dassonvillei subsp. dassonvillei DSM 43235 (wild-type producer of Protease 18), Nocardiopsis prasina DSM 15648 (wild-type producer of Protease 11), Nocardiopsis prasina DSM 15649 (wild-type producer of Protease 35), Nocardiopsis sp. NRRL 18262 (wild-type producer of Protease 10), Nocardiopsis sp. FERM P-18676 (described in JP 2003284571-A).

Strains of these species are accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL), e.g. Nocardiopsis dassonvillei subsp. dassonvillei DSM 43235 is publicly available from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany).

Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms or DNA isolated from nature (e.g., soil, composts, water, etc.) using suitable probes. Techniques for isolating microorganisms or DNA from natural habitats are well known in the art. The nucleic acid sequence may then be derived by similarly screening a genomic or cDNA library of another microorganism. Once a nucleic acid sequence encoding a polypeptide has been detected with the probe(s), the sequence may be isolated or cloned by utilizing techniques which are known to those of ordinary skill in the art (see, e.g., Sambrook et al., 1989, supra).

The parent protease may be a mature part of any of the amino acid sequences referred to above. A mature part means a mature amino acid sequence and refers to that part of an amino acid sequence which remains after a potential signal peptide part and/or pro-peptide part has been cleaved off. The mature parts of each of the proteases Protease 08, 10, 11, 18, 22 and 35 are specified in the appended sequence listing.

The parent protease may also be a fragment of a specified amino acid sequence, viz. a polypeptide having one or more amino acids deleted from the amino and/or carboxyl terminus of this amino acid sequence. In one embodiment, a fragment contains at least 80, or at least 90, or at least 100, or at least 110, or at least 120, or at least 130, or at least 140, or at least 150, or at least 160, or at least 170, or at least 180, or at least 185 amino acid residues.

The parent protease may also be an allelic variant, allelic referring to the existence of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

In another embodiment, the parent protease may be a genetically engineered protease, e.g. a variant of the wild-type or natural parent proteases referred to above comprising a substitution, deletion, and/or insertion of one or more amino acids. In other words: The parent protease may itself be a protease variant, such as Protease 22. The amino acid sequence of such parent protease may differ from the amino acid sequence specified by an insertion or deletion of one or more amino acid residues and/or the substitution of one or more amino acid residues by different amino acid residues. The amino acid changes may be of a minor, or of a major, nature. Amino acid changes of a major nature are e.g. those resulting in a variant protease of the present invention with amended properties. In another particular embodiment, the amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter the specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, In, The Proteins, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly as well as these in reverse.

Still further examples of genetically engineered parent proteases are synthetic

proteases, designed by man, and expectedly not occurring in nature. EP 897985 discloses a process of preparing a consensus protein. Shuffled proteases are other examples of synthetic or genetically engineered parent proteases, which can be prepared as is generally known in the art, eg by Site-directed Mutagenesis, by PCR (using a PCR fragment containing the desired mutation as one of the primers in the PCR reactions), or by Random Mutagenesis. Included in the concept of a synthetic protease is also any hybrid or chimeric protease, i.e. a protease which comprises a combination of partial amino acid sequences derived from at least two proteases. Gene shuffling is generally described in e.g. WO 95/22625 and WO 96/00343. Re-combination of protease genes can be made independently of the specific sequence of the parents by synthetic shuffling as described in Ness, J.E. et al, in Nature Biotechnology, Vol. 20 (12), pp. 1251-1255, 2002. Synthetic oligonucleotides degenerated in their DNA sequence to provide the possibility of all amino acids found in the set of parent proteases are designed and the genes assembled according to the reference. The shuffling can be carried out for the full length sequence or for only part of the sequence and then later combined with the rest of the gene to give a full length sequence. Two, three, four, five or all six of the the proteases designated Protease 10, 18, 11, 35, 08 and 22 (SEQ ID NOs: 2, 4, 6, 8, 10, and 21; in particular the mature parts thereof) are particular examples of such parent proteases which can be subjected to shuffling as described above, to provide additional proteases of the invention.

In further particular embodiments, the parent protease comprises, or consists of, respectively, the amino acid sequence specified, or an allelic variant thereof; or a fragment thereof that has protease activity.

In still further particular embodiments, the protease variant of the invention is not identical to: (i) amino acids 1-188 of SEQ ID NO: 2, amino acids 1-188 of SEQ ID NO: 4, amino acids 1-188 of SEQ ID NO: 6, amino acids 1-188 of SEQ ID NO: 8, and amino acids 1-188 of SEQ ID NO: 10; (ii) amino acids 1-188 of SEQ ID NO: 2; (iii) amino acids 1-188 of SEQ ID NO: 2 with the substitution T87A; (iv) amino acids 1-188 of SEQ ID NO: 4; (v) amino acids 1-188 of SEQ ID NO: 6; (vi) amino acids 1-188 of SEQ ID NO: 8; (vii) amino acids 1-188 of SEQ ID NO: 10; (viii) the protease derived from Nocardiopsis dassonvillei NRRL 18133; (ix) the protease having amino acids 1 to 188 of SEQ ID NO: 2 as disclosed in JP 2003284571-A; (x) the protease having the sequence entered in GENESEQP with no. ADF43564; (xi) the protease disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 2, in particular the mature part thereof; (xii) the protease disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 4, in particular the mature part thereof; (xiii) the protease disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 6, in particular the mature part thereof; (xiv) the protease disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 8, in particular the mature part thereof; (xv) the protease disclosed in DK patent application no. 2004 00969 as

SEQ ID NO: 10, in particular the mature part thereof; (xvi) the protease disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 12, in particular the mature part thereof; and/or (xvii) any prior art protease of a percentage of identity to SEQ ID NO: 2 of at least 60%.

Microorganism Taxonomy

Questions relating to taxonomy may be solved by consulting a taxonomy data base, such as the NCBI Taxonomy Browser which is available at the following internet site: http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/, and/or by consulting Taxonomy handbooks. For the present purposes, the taxonomy is preferably according to the chapter: The road map to the Manual by G.M. Garrity & J. G. Holt in Bergey's Manual of Systematic Bacteriology, 2001, second edition, volume 1, David R. Bone, Richard W. Castenholz.

Amino Acid Homology

The present invention refers to proteases, viz. parent proteases, and/or protease variants, having a certain degree of identity to amino acids 1 to 188 of SEQ ID NO: 2, such parent and/or variant proteases being hereinafter designated "homologous proteases".

For purposes of the present invention the degree of identity between two amino acid sequences, as well as the degree of identity between two nucleotide sequences, is determined by the program "align" which is a Needleman-Wunsch alignment (i.e. a global alignment). The program is used for alignment of polypeptide, as well as nucleotide sequences. The default scoring matrix BLOSUM50 is used for polypeptide alignments, and the default identity matrix is used for nucleotide alignments. The penalty for the first residue of a gap is -12 for polypeptides and -16 for nucleotides. The penalties for further residues of a gap are -2 for polypeptides, and -4 for nucleotides.

"Align" is part of the FASTA package version v20u6 (see W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448, and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA," Methods in Enzymology 183:63-98). FASTA protein alignments use the Smith-Waterman algorithm with no limitation on gap size (see "Smith-Waterman algorithm", T. F. Smith and M. S. Waterman (1981) J. Mol. Biol. 147:195-197).

Multiple alignments of protein sequences may be made using "ClustalW" (Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680). Multiple alignment of DNA sequences may be done using the protein alignment as a template, replacing the amino acids with the corresponding codon from the DNA sequence.

In particular embodiments, the homologous protease has an amino acid sequence

which has a degree of identity to amino acids 1 to 188 of SEQ ID NO: 2 of at least 60%, 62%, 64%, 66%, 68%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or of at least about 99%.

In alternative embodiments, the homologous protease has an amino acid sequence which has a degree of identity to SEQ ID NO: 2 of at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, or at least 59%.

In another particular embodiment, the parent protease, and/or the protease variant, comprises a mature amino acid sequence which differs by no more than seventyfive, seventyfour, seventythree, seventytwo, seventyone, seventy, sixtynine, sixtyeight, sixtyseven, sixtysix, sixtyfive, sixtyfour, sixtythree, sixtytwo, sixtyone, sixty, fiftynine, fiftyeight, fiftyseven, fiftysix, fiftyfive, fiftyfour, fiftythree, fiftytwo, fiftyone, fifty, fortynine, fortyeight, fortyseven, fortysix, fortyfive, fortyfour, fortythree, fortytwo, fortyone, forty, thirtynine, thirtyeight, thirtyseven, thirtysix, thirtyfive, thirtyfour, thirtythree, thirtytwo, thirtyone, thirty, twentynine, twentyeight, twentyseven, twentysix, twentyfive, twentyfour, twentythree, twentytwo, twentyone, twenty, nineteen, eighteen, seventeen, sixteen, fifteen, fourteen, thirteen, twelve, eleven, ten, nine, eight, seven, six, five, four, three, by no more than two, or only by one amino acid(s) from the specified amino acid sequence, e.g. amino acids 1 to 188 of SEQ ID NO: 2.

In a still further particular embodiment, the parent protease, and/or the protease variant, comprises a mature amino acid sequence which differs by at least seventyfive, seventyfour, seventythree, seventytwo, seventyone, seventy, sixtynine, sixtyeight, sixtyseven, sixtysix, sixtyfive, sixtyfour, sixtythree, sixtytwo, sixtyone, sixty, fiftynine, fiftyeight, fiftyseven, fiftysix, fiftyfive, fiftyfour, fiftythree, fiftytwo, fiftyone, fifty, fortynine, fortyeight, fortyseven, fortysix, fortyfive, fortyfour, fortythree, fortytwo, fortyone, forty, thirtynine, thirtyeight, thirtyseven, thirtysix, thirtyfive, thirtyfour, thirtythree, thirtytwo, thirtyone, thirty, twentynine, twentyeight, twentyseven, twentysix, twentyfive, twentyfour, twentythree, twentytwo, twentyone, twenty, nineteen, eighteen, seventeen, sixteen, fifteen, fourteen, thirteen, twelve, eleven, ten, nine, eight, seven, six, five, four, three, by at least two, or by one amino acid(s) from the specified amino acid sequence, e.g. amino acids 1 to 188 of SEQ ID NO: 2.

Nucleic Acid Hybridization

In the alternative, homologous parent proteases, as well as variant proteases, may be defined as being encoded by a nucleic acid sequence which hybridizes under very low stringency conditions, preferably low stringency conditions, more preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with nucleotides 900-1466, or 900-1463, of SEQ ID NO: 1, or a subsequence or a complementary

strand thereof (J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, New York). A subsequence may be at least 100 nucleotides, or at least 200, 300, 400, or at least 500 nucleotides. Moreover, the subsequence may encode a polypeptide fragment that has the relevant enzyme activity.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2 x SSC, 0.2% SDS preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

For short probes which are about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at 5°C to 10°C below the calculated T_m using the calculation according to Bolton and McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures.

For short probes which are about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SCC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated T_m.

Position numbering

In the present context, the basis for numbering positions is amino acids 1 to 188 of SEQ ID NO: 2, Protease 10, starting with A1 and ending with T188, see Fig. 1. A parent protease, as well as a variant protease, may comprise extensions as compared to SEQ ID NO: 2, i.e. in the N-terminal, and/or the C-terminal ends thereof. The amino acids of such extensions, if any, are to be numbered as is usual in the art, i.e. for a C-terminal extension: 189, 190, 191 and so forth, and for an N-terminal extension -1, -2, -3 and so forth.

Alterations, such as Substitutions, Deletions, Insertions

In the present context, the following are examples of various ways in which a protease

variant can be designed or derived from a parent amino acid sequence: An amino acid can be substituted with another amino acid; an amino acid can be deleted; an amino acid can be inserted; as well as any combination of any number of such alterations.

For the present purposes, the term substitution is intended to include any number of any type of such alterations. This is a reasonable definition, because, for example, a deletion can be regarded as a substitution of an amino acid, AA, in a given position, nn, with nothing, (). Such substitution can be designated: AAnn(). Likewise, an insertion of only one amino acid, BB, downstream an amino acid, AA, in a given position, nn, can be designated: ()nnaBB. And if two amino acids, BB and CC, are inserted downstream of amino acid AA in position nn, this substitution (combination of two substitutions) can be designated: ()nnaBB+()nnbCC, the thus created gaps between amino acids nn and nn+1 in the parent sequence being assigned lower case or subscript letters a, b, c etc. to the former position number, here nn. A similar numbering procedure is followed when aligning a new sequence to the multiple alignment of Fig. 1, in case of a gap being created by the alignment between amino acids nn and nn+1: Each position of the gap is assigned a number: nna, nnb etc.. A comma (,) between substituents, as e.g. in the substitution T129E,D,Y,Q means "either or", i.e. that T129 is substituted with E, or D, or Y, or Q. A plus-sign (+) between substitutions, e.g. 129D+135P means "and", i.e. that these two single substitutions are combined in one and the same protease variant.

In the present context, the term "a" substitution" means at least one substitution. At least one means one or more, e.g. one, or two, or three, or four, or five, or six, or seven, or eight, or nine, or ten, or twelve, or fourteen, or fifteen, or sixteen, or eighteen, or twenty, or twentytwo or twentyfour, or twentyfive, or twenty eight, or thirty, and so on, to include in principle, any number of substitutions. The variants of the invention, however, still have to be, e.g., at least 60% identical to SEQ ID NO: 2, this percentage being determined by the above-mentioned program. The substitutions can be applied to any position encompassed by any region mentioned in claim 1, and variants comprising combinations of any number and type of such substitutions are also included. The term substitution as used herein also include deletions, as well as extensions, or insertions, that may add to the length of the sequence corresponding to amino acids 1 to 188 of SEQ ID NO: 2.

Furthermore, the term "a substitution" embraces a substitution into any one of the other nineteen natural amino acids, or into other amino acids, such as non-natural amino acids. For example, a substitution of amino acid T in position 22 includes each of the following substitutions: 22A, 22C, 22D, 22E, 22F, 22G, 22H, 22I, 22K, 22L, 22M, 22N, 22P, 22Q, 22R, 22S, 22V, 22W, and 22Y. This is, by the way, equivalent to the designation 22X, wherein X designates any amino acid. These substitutions can also be designated T22A, T22C, T22X, etc. The same applies by analogy to each and every position mentioned herein, to specifically

include herein any one of such substitutions.

Identifying Corresponding Position Numbers

For each amino acid residue in each parent, or variant, protease of the invention, and/or for use according to the invention, it is possible to directly and unambiguously assign an amino acid residue in the sequence of amino acids 1 to 188 of SEQ ID NO: 2 to which it corresponds. Corresponding residues are assigned the same number, by reference to the Protease 10 sequence.

As it appears from the numbering of Fig. 1, in conjunction with the numbering of the sequence listing, for each amino acid residue of each of the proteases Protease 10, Protease 18, Protease 11, Protease 35, Protease 08, and Protease 22, the corresponding amino acid residue in SEQ ID NO: 2 has the same number. This number is easily derivable from Fig. 1. At least in case of these six proteases, the number is the same as the number assigned to this amino acid residue in the sequence listing for the mature part of the respective protease.

For a given position in another protease - be it a parent or a variant protease - a corresponding position of SEQ ID NO: 2 can always be found, as follows:

The amino acid sequence of another parent protease, or, in turn, of a variant protease amino acid sequence, is designated SEQ-X. A position corresponding to position N of SEQ ID NO: 2 is found as follows: The parent or variant protease amino acid sequence SEQ-X is aligned with SEQ ID NO: 2 as specified above in the section entitled Amino Acid Homology. From the alignment, the position in sequence SEQ-X corresponding to position N of SEQ ID NO: 2 can be clearly and unambiguously derived, using the principles described below.

SEQ-X is the mature part of the protease in question. In the alternative, it may also include a signal peptide part, and/or a propeptide part, or it may be a fragment of the mature protease which has protease activity, e.g. a fragment of the same length as SEQ ID NO: 2, and/or it may be the fragment which extends from A1 to T188 when aligned with SEQ ID NO: 2 as described herein.

Region and Position

In the present context, the term region means at least one position of a parent protease amino acid sequence, the term position designating an amino acid residue of such amino acid sequence. In one embodiment, region means one or more successive positions of the parent protease amino acid sequence, e.g. one, two, three, four, five, six, seven, eight, etc., up to any number of consecutive positions of the sequence. Accordingly, a region may consist of one position only, or it may consist of any number of consecutive positions, such as, e.g., position no. 62 and 63; or position no. 111, 112, 113 and 114. For the present purposes, these two regions are designated 62-63, and 111-114, respectively. The boundaries of these

regions or ranges are included in the region.

A region encompasses specifically each and every position it embraces. For example, region 111-114 specifically encompasses each of the positions 111, 112, 113, and 114. The same applies by analogy for the other regions mentioned herein.

Thermostability

For the present purposes, the term thermostable as applied in the context of a certain polypeptide, refers to the melting temperature, Tm, of such polypeptide, as determined using Differential Scanning Calorimetry (DSC) in 10mM sodium phosphate, 50 mM sodium chloride, pH 7.0, using a constant scan rate of 1.5°C/min.

The following Tm's were determined under the above conditions: 76.5°C (Protease 10), 83.0°C (Protease 18), 78.3°C (Protease 08), 76.6°C (Protease 35), 73.7°C (Protease 11), and 83.5°C (Protease 22).

For a thermostable polypeptide, the Tm is at least 83.1°C. In particular embodiments, the Tm is at least 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or at least 100°C.

In the alternative, the term thermostable refers to a melting temperature of at least 73.8, or at least 76.7°C, or at least 78.4°C, preferably at least 74, 75, 76, 77, 78, 79, 80, 81, 82, or at least 83°C, still as determined using DSC at a pH of 7.0.

For the determination of Tm, a sample of the polypeptide with a purity of at least 90% (or 91, 92, 93, 94, 95, 96, 97, or 98%) as determined by SDS-PAGE may be used. Still further, the enzyme sample may have a concentration of between 0.5 and 2.5 mg/ml protein (or between 0.6 and 2.4, or between 0.7 and 2.2, or between 0.8 and 2.0 mg/ml protein), as determined from absorbance at 280 nm and based on an extinction coefficient calculated from the amino acid sequence of the enzyme in question.

The DSC takes place at the desired pH (e.g. pH 5.5, 7.0, 3.0, or 2.5) and with a constant heating rate, e.g. of 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9 or 10°C/min.

In a particular embodiment, the protease variant of the invention is thermostable, preferably more thermostable than the parent protease. In this context, preferred parent proteases are Protease 18, or Protease 10.

In another particular embodiment, a culture supernatant of the protease variant of the invention, appropriately diluted, exhibits a residual activity after incubation for four hours at 65°C in a 0.2M Na_2HPO_4 buffer, titrated with 0.1M citric acid to i) pH 6.0, or ii) pH 4.0, of at least 20%, relative to an un-incubated (frozen) control, the activity being measured using the Protazyme AK assay at pH 8.5 and 37°C, as described in Example 2. In further particular embodiments, the residual activity is at least 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or at least 77%.

Temperature Activity Profile

In a particular embodiment, the protease variant of the invention exhibits an amended temperature activity profile as compared to, e.g., Protease 10 (or Protease 18, Protease 11, Protease 35, or Protease 08). For example, the protease variant of the invention may exhibit a relative activity at pH 9 and 80°C of at least 0.40, preferably at least 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, or at least 0.95, the term "relative" referring to the maximum activity measured for the protease in question. For Protease 22, the activity is relative to the activity at 80°C which is set to 1.000 (100%), and for Protease 10, the activity at 70°C is set to 1.000 (100%), see Example 3. As another example, the protease variant of the invention exhibits a relative activity at pH 9 and 90°C of at least 0.10, preferably at least 0.15, 0.20, 0.25, 0.30, or of at least 0.35. In a particular embodiment, the protease activity is measured using the Protazyme AK assay of Example 1.

Low-allergenic Variants

In a specific embodiment, the protease variants of the present invention are (also) low-allergenic variants, designed to invoke a reduced immunological response when exposed to animals, including man. The term immunological response is to be understood as any reaction by the immune system of an animal exposed to the protease variant. One type of immunological response is an allergic response leading to increased levels of IgE in the exposed animal. Low-allergenic variants may be prepared using techniques known in the art. For example the protease variant may be conjugated with polymer moieties shielding portions or epitopes of the protease variant involved in an immunological response. Conjugation with polymers may involve in vitro chemical coupling of polymer to the protease variant, e.g. as described in WO 96/17929, WO 98/30682, WO 98/35026, and/or WO 99/00489. Conjugation may in addition or alternatively thereto involve in vivo coupling of polymers to the protease variant. Such conjugation may be achieved by genetic engineering of the nucleotide sequence encoding the protease variant, inserting consensus sequences encoding additional glycosylation sites in the protease variant and expressing the protease variant in a host capable of glycosylating the protease variant, see e.g. WO 00/26354. Another way of providing low-allergenic variants is genetic engineering of the nucleotide sequence encoding the protease variant so as to cause the protease variants to self-oligomerize, effecting that protease variant monomers may shield the epitopes of other protease variant monomers and thereby lowering the antigenicity of the oligomers. Such products and their preparation is described e.g. in WO 96/16177. Epitopes involved in an immunological response may be identified by various methods such as the phage display method described in WO 00/26230 and WO 01/83559, or the random approach described in EP 561907. Once an epitope has

been identified, its amino acid sequence may be altered to produce altered immunological properties of the protease variant by known gene manipulation techniques such as site directed mutagenesis (see e.g. WO 00/26230, WO 00/26354 and/or WO 00/22103) and/or conjugation of a polymer may be done in sufficient proximity to the epitope for the polymer to shield the epitope.

Nucleic Acid Sequences and Constructs

The present invention also relates to nucleic acid sequences comprising a nucleic acid sequence which encodes a protease variant of the invention.

The term "isolated nucleic acid sequence" refers to a nucleic acid sequence which is essentially free of other nucleic acid sequences, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably at least about 60% pure, even more preferably at least about 80% pure, and most preferably at least about 90% pure as determined by agarose electrophoresis. For example, an isolated nucleic acid sequence can be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the nucleic acid sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host cell where multiple copies or clones of the nucleic acid sequence will be replicated. The nucleic acid sequence may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

The nucleic acid sequences of the invention can be prepared by introducing at least one mutation into the parent protease coding sequence or a subsequence thereof, wherein the mutant nucleic acid sequence encodes a variant protease. The introduction of a mutation into the nucleic acid sequence to exchange one nucleotide for another nucleotide may be accomplished by site-directed mutagenesis using any of the methods known in the art, e.g. by site-directed mutagenesis, by random mutagenesis, or by doped, spiked, or localized random mutagenesis.

Random mutagenesis is suitably performed either as localized or region-specific random mutagenesis in at least three parts of the gene translating to the amino acid sequence shown in question, or within the whole gene. When the mutagenesis is performed by the use of an oligonucleotide, the oligonucleotide may be doped or spiked with the three non-parent nucleotides during the synthesis of the oligonucleotide at the positions which are to be changed. The doping or spiking may be performed so that codons for unwanted amino acids are avoided. The doped or spiked oligonucleotide can be incorporated into the DNA encoding the protease enzyme by any technique, using, e.g., PCR, LCR or any DNA polymerase and ligase as deemed appropriate.

Preferably, the doping is carried out using "constant random doping", in which the percentage of wild-type and mutation in each position is predefined. Furthermore, the doping may be directed toward a preference for the introduction of certain nucleotides, and thereby a preference for the introduction of one or more specific amino acid residues. The doping may be made, e.g., so as to allow for the introduction of 90% wild type and 10% mutations in each position. An additional consideration in the choice of a doping scheme is based on genetic as well as protein-structural constraints.

The random mutagenesis may be advantageously localized to a part of the parent protease in question. This may, e.g., be advantageous when certain regions of the enzyme have been identified to be of particular importance for a given property of the enzyme.

Alternative methods for providing variants of the invention include gene shuffling e.g. as described in WO 95/22625 or in WO 96/00343, and the consensus derivation process as described in EP 897985 (see the section "Parent Protease" for more details).

In particular embodiments, the nucleic acid sequence of the invention is not identical to: (i) Nucleotides 900-1466, or 900-1463, of SEQ ID NO: 1, nucleotides 499-1062 of SEQ ID NO: 3, nucleotides 496-1059 of SEQ ID NO: 5, nucleotides 496-1059 of SEQ ID NO: 7, and nucleotides 502-1065 of SEQ ID NO: 9; (ii) nucleotides 900-1466 of SEQ ID NO: 1; (iii) nucleotides 900-1463 of SEQ ID NO: 1; (iv) nucleotides 900-1463 of SEQ ID NO: 1 as disclosed in DK 1996 00013; (v) nucleotides 499-1062 of SEQ ID NO: 3; (vi) nucleotides 496-1059 of SEQ ID NO: 5; (vii) nucleotides 496-1059 of SEQ ID NO: 7; (viii) nucleotides 502-1065 of SEQ ID NO: 9; (xi) the nucleic acid sequence encoding the mature peptide part of the protease derived from Nocardiopsis dassonvillei NRRL 18133; (x) the nucleic acid sequence having SEQ ID NO: 1 as disclosed in JP 2003284571-A; (xi) the nucleic acid sequence GENESEQN no. ADF43563; (xii) the nucleic acid sequence disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 1, in particular the mature peptide encoding part thereof; (xiii) the nucleic acid sequencep disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 3, in particular the mature peptide encoding part thereof; (xiv) the nucleic acid sequence disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 5, in particular the mature peptide encoding part thereof; (xv) the nucleic acid sequence disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 7, in particular the mature peptide encoding part thereof; (xvi) the nucleic acid sequence disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 9, in particular the mature peptide encoding part thereof; (xvii) the nucleic acid sequence disclosed in DK patent application no. 2004 00969 as SEQ ID NO: 11, in particular the mature peptide encoding part thereof; and/or (xviii) nucleic acid sequences encoding any prior art proteases of at least 60% identity to amino acids 1 to 188 of SEQ ID NO: 2.

Nucleic Acid Constructs

A nucleic acid construct comprises a nucleic acid sequence of the present invention operably linked to one or more control sequences which direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences. Expression will be understood to include any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

Expression vector

A nucleic acid sequence encoding a protease variant of the invention can be expressed using an expression vector which typically includes control sequences encoding a promoter, operator, ribosome binding site, translation initiation signal, and, optionally, a repressor gene or various activator genes.

The recombinant expression vector carrying the DNA sequence encoding a protease variant of the invention may be any vector which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. The vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The protease variant may also be co-expressed together with at least one other enzyme of animal feed interest, such as an alpha-amylase, a phytase, a galactanase, a xylanase, an endoglucanase, an endo-1,3(4)-beta-glucanase, an alpha-galactosidase, and/or a protease. The enzymes may be co-expressed from different vectors, from one vector, or using a mixture of both techniques. When using different vectors, the vectors may have different selectable markers, and different origins of replication. When using only one vector, the genes can be expressed from one or more promoters. If cloned under the regulation of one promoter (di- or multi-cistronic), the order in which the genes are cloned may affect the expression levels of the proteins. The protease variant may also be expressed as a fusion protein, i.e. that the gene encoding the protease variant has been fused in frame to the gene encoding another protein. This protein may be another enzyme or a functional domain from another enzyme.

Host Cells

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. A vector comprising a nucleic acid sequence of the present invention is introduced into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector. The term "host cell" encompasses any progeny of a

parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source.

The host cell may be a unicellular microorganism, e.g., a prokaryote, or a non-unicellular microorganism, e.g., a eukaryote cell, such as an animal, a mammalian, an insect, a plant, or a fungal cell. Preferred animal cells are non-human animal cells.

In a preferred embodiment, the host cell is a fungal cell, or a yeast cell, such as a Candida, Hansenula, Kluyveromyces, Pichia, Saccharomyces, Schizosaccharomyces, or Yarrowia cell. The fungal host cell may be a filamentous fungal cell, such as a cell of a species of, but not limited to, Acremonium, Aspergillus, Fusarium, Humicola, Mucor, Myceliophthora, Neurospora, Penicillium, Thielavia, Tolypocladium, or Trichoderma. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a Bacillus cell, e.g., Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus brevis, Bacillus circulans, Bacillus clausii, Bacillus coagulans, Bacillus lautus, Bacillus lentus, Bacillus licheniformis, Bacillus megaterium, Bacillus stearothermophilus, Bacillus subtilis, and Bacillus thuringiensis, or a Streptomyces cell, such as Streptomyces lividans or Streptomyces murinus, or a Nocardiopsis cell, or cells of lactic acid bacteria; or gram negative bacteria such as E. coli and Pseudomonas sp. Lactic acid bacteria include, but are not limited to, species of the genera Lactococcus, Lactobacillus, Leuconostoc, Streptococcus, Pediococcus, and Enterococcus.

Methods of Production

The present invention also relates to methods for producing a protease variant of the present invention comprising (a) cultivating a host cell under conditions conducive for production of the protease variant; and (b) recovering the protease variant.

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the protease is secreted into the nutrient medium, it can be recovered directly from the medium. If it is not secreted, it can be recovered from cell lysates.

The resulting protease may be recovered by methods known in the art. For example, it can be recovered from the nutrient medium by conventional procedures including, but not

limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

The proteases of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulfate precipitation), SDS-PAGE, or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Plants

The present invention also relates to a transgenic plant, plant part, or plant cell which has been transformed with a nucleic acid sequence encoding a polypeptide having protease activity of the present invention so as to express and produce the polypeptide in recoverable quantities. The polypeptide may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide may be used as such for improving the quality of a food or feed, e.g., improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive factor.

In a particular embodiment, the polypeptide is targeted to the endosperm storage vacuoles in seeds. This can be obtained by synthesizing it as a precursor with a suitable signal peptide, see Horvath et al in PNAS, Feb. 15, 2000, vol. 97, no. 4, p. 1914-1919.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot) or engineered variants thereof. Examples of monocot plants are grasses, such as meadow grass (blue grass, Poa), forage grass such as Festuca, Lolium, temperate grass, such as Agrostis, and cereals, e.g., wheat, oats, rye, barley, rice, sorghum, triticale (stabilized hybrid of wheat (Triticum) and rye (Secale), and maize (corn). Examples of dicot plants are tobacco, legumes, such as sunflower (Helianthus), cotton (Gossypium), lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family Brassicaceae), such as cauliflower, rape seed, and the closely related model organism Arabidopsis thaliana. Low-phytate plants as described e.g. in US patent no. 5,689,054 and US patent no. 6,111,168 are examples of engineered plants.

Examples of dicot plants are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family Brassicaceae), such as cauliflower, rape seed, and the closely related model organism Arabidopsis thaliana. Low-phytate plants as described e.g. in US patent no. 5,689,054 and US patent no. 6,111,168 are examples of engineered plants. Examples of plant parts are stem, callus, leaves, root, fruits, seeds, and tubers, as well as the individual tissues comprising these parts, e.g. epidermis, mesophyll, parenchyma, vascular tissues, meristems. Also specific plant cell compartments, such as chloroplast, apoplast, mitochondria, vacuole, peroxisomes, and cytoplasm are considered to

be a plant part. Furthermore, any plant cell, whatever the tissue origin, is considered to be a plant part. Likewise, plant parts such as specific tissues and cells isolated to facilitate the utilisation of the invention are also considered plant parts, e.g. embryos, endosperms, aleurone and seed coats.

Also included within the scope of the present invention are the progeny of such plants, plant parts and plant cells.

The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with methods known in the art. Briefly, the plant or plant cell is constructed by incorporating one or more expression constructs encoding a polypeptide of the present invention into the plant host genome and propagating the resulting modified plant or plant cell into a transgenic plant or plant cell.

Conveniently, the expression construct is a nucleic acid construct which comprises a nucleic acid sequence encoding a polypeptide of the present invention operably linked with appropriate regulatory sequences required for expression of the nucleic acid sequence in the plant or plant part of choice. Furthermore, the expression construct may comprise a selectable marker useful for identifying host cells into which the expression construct has been integrated and DNA sequences necessary for introduction of the construct into the plant in question (the latter depends on the DNA introduction method to be used).

The choice of regulatory sequences, such as promoter and terminator sequences and optionally signal or transit sequences are determined, for example, on the basis of when, where, and how the polypeptide is desired to be expressed. For instance, the expression of the gene encoding a polypeptide of the present invention may be constitutive or inducible, or may be developmental, stage or tissue specific, and the gene product may be targeted to a specific tissue or plant part such as seeds or leaves. Regulatory sequences are, for example, described by Tague et al., 1988, Plant Physiology 86: 506.

For constitutive expression, the following promoters may be used: The 35S-CaMV promoter (Franck et al., 1980, Cell 21: 285-294), the maize ubiquitin 1 (Christensen AH, Sharrock RA and Quail 1992. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation), or the rice actin 1 promoter (Plant Mo. Biol. 18, 675-689.; Zhang W, McElroy D. and Wu R 1991, Analysis of rice Act1 5' region activity in transgenic rice plants. Plant Cell 3, 1155-1165). Organ-specific promoters may be, for example, a promoter from storage sink tissues such as seeds, potato tubers, and fruits (Edwards & Coruzzi, 1990, Ann. Rev. Genet. 24: 275-303), or from metabolic sink tissues such as meristems (Ito et al., 1994, Plant Mol. Biol. 24: 863-878), a seed specific promoter such as the glutelin, prolamin, globulin, or albumin promoter from rice (Wu et al., 1998, Plant and Cell Physiology 39: 885-889), a Vicia faba promoter from the legumin B4 and the unknown seed protein gene from Vicia faba (Conrad et

al., 1998, Journal of Plant Physiology 152: 708-711), a promoter from a seed oil body protein (Chen et al., 1998, Plant and Cell Physiology 39: 935-941), the storage protein napA promoter from Brassica napus, or any other seed specific promoter known in the art, e.g., as described in WO 91/14772. Furthermore, the promoter may be a leaf specific promoter such as the rbcs promoter from rice or tomato (Kyozuka et al., 1993, Plant Physiology 102: 991-1000, the chlorella virus adenine methyltransferase gene promoter (Mitra and Higgins, 1994, Plant Molecular Biology 26: 85-93), or the aldP gene promoter from rice (Kagaya et al., 1995, Molecular and General Genetics 248: 668-674), or a wound inducible promoter such as the potato pin2 promoter (Xu et al., 1993, Plant Molecular Biology 22: 573-588). Likewise, the promoter may be inducible by abiotic treatments such as temperature, drought or alterations in salinity or inducible by exogenously applied substances that activate the promoter, e.g. ethanol, oestrogens, plant hormones like ethylene, abscisic acid, gibberellic acid, and/or heavy metals.

A promoter enhancer element may also be used to achieve higher expression of the enzyme in the plant. For instance, the promoter enhancer element may be an intron which is placed between the promoter and the nucleotide sequence encoding a polypeptide of the present invention. For instance, Xu et al., 1993, supra disclose the use of the first intron of the rice actin 1 gene to enhance expression.

Still further, the codon usage may be optimized for the plant species in question to improve expression (see Horvath et al referred to above).

The selectable marker gene and any other parts of the expression construct may be chosen from those available in the art.

The nucleic acid construct is incorporated into the plant genome according to conventional techniques known in the art, including Agrobacterium-mediated transformation, virus-mediated transformation, microinjection, particle bombardment, biolistic transformation, and electroporation (Gasser et al., 1990, Science 244: 1293; Potrykus, 1990, Bio/Technology 8: 535; Shimamoto et al., 1989, Nature 338: 274).

Presently, Agrobacterium tumefaciens-mediated gene transfer is the method of choice for generating transgenic dicots (for a review, see Hooykas and Schilperoort, 1992, Plant Molecular Biology 19: 15-38), and it can also be used for transforming monocots, although other transformation methods are generally preferred for these plants. Presently, the method of choice for generating transgenic monocots, supplementing the Agrobacterium approach, is particle bombardment (microscopic gold or tungsten particles coated with the transforming DNA) of embryonic calli or developing embryos (Christou, 1992, Plant Journal 2: 275-281; Shimamoto, 1994, Current Opinion Biotechnology 5: 158-162; Vasil et al., 1992, Bio/Technology 10: 667-674). An alternative method for transformation of monocots is based on protoplast transformation as described by Omirulleh et al., 1993, Plant Molecular Biology

21: 415-428.

Following transformation, the transformants having incorporated therein the expression construct are selected and regenerated into whole plants according to methods well-known in the art.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a transgenic plant or a plant cell comprising a nucleic acid sequence encoding a protease variant of the present invention under conditions conducive for production of the protease variant; and (b) recovering the protease variant.

Animals as Expression Hosts

The present invention also relates to a transgenic, non-human animal and products or elements thereof, examples of which are body fluids such as milk and blood, organs, flesh, and animal cells. Techniques for expressing proteins, e.g. in mammalian cells, are known in the art, see e.g. the handbook Protein Expression: A Practical Approach, Higgins and Hames (eds), Oxford University Press (1999), and the three other handbooks in this series relating to Gene Transcription, RNA processing, and Post-translational Processing. Generally speaking, to prepare a transgenic animal, selected cells of a selected animal are transformed with a nucleic acid sequence encoding a protease variant of the present invention so as to express and produce the protease variant. The protease variant may be recovered from the animal, e.g. from the milk of female animals, or it may be expressed to the benefit of the animal itself, e.g. to assist the animal's digestion. Examples of animals are mentioned below in the section headed Animal Feed and Animal Feed Additives.

To produce a transgenic animal with a view to recovering the protease variant from the milk of the animal, a gene encoding the protease variant may be inserted into the fertilized eggs of an animal in question, e.g. by use of a transgene expression vector which comprises a suitable milk protein promoter, and the gene encoding the protease variant. The transgene expression vector is microinjected into fertilized eggs, and preferably permanently integrated into the chromosome. Once the egg begins to grow and divide, the potential embryo is implanted into a surrogate mother, and animals carrying the transgene are identified. The resulting animal can then be multiplied by conventional breeding. The protease variant may be purified from the animal's milk, see e.g. Meade, H.M. et al (1999): Expression of recombinant proteins in the milk of transgenic animals, Gene expression systems: Using nature for the art of expression. J. M. Fernandez and J. P. Hoeffler (eds.), Academic Press.

In the alternative, in order to produce a transgenic non-human animal that carries in the genome of its somatic and/or germ cells a nucleic acid sequence including a heterologous transgene construct including a transgene encoding the protease variant, the transgene may be operably linked to a first regulatory sequence for salivary gland specific expression of the

protease variant, as disclosed in WO 2000064247.

Animal Feed and Animal Feed Additives

For the present purposes, the term animal includes all animals, including human beings. In a particular embodiment, the protease variants and compositions of the invention can be used as a feed additive for non-human animals. Examples of animals are non-ruminants, and ruminants, such as sheep, goats, horses, and cattle, e.g. beef cattle, cows, and young calves. In a particular embodiment, the animal is a non-ruminant animal. Non-ruminant animals include mono-gastric animals, e.g. pigs or swine (including, but not limited to, piglets, growing pigs, and sows); poultry such as turkeys, ducks and chicken (including but not limited to broiler chicks, layers); young calves; and fish (including but not limited to salmon, trout, tilapia, catfish and carps; and crustaceans (including but not limited to shrimps and prawns).

The term feed or feed composition means any compound, preparation, mixture, or composition suitable for, or intended for intake by an animal. The feed can be fed to the animal before, after, or simultaneously with the diet. The latter is preferred.

The composition of the invention, when intended for addition to animal feed, may be designated an animal feed additive. Such additive always comprises the protease variant in question, preferably in the form of stabilized liquid or dry compositions. The additive may comprise other components or ingredients of animal feed. The so-called pre-mixes for animal feed are particular examples of such animal feed additives. Pre-mixes may contain the enzyme(s) in question, and in addition at least one vitamin and/or at least one mineral.

Accordingly, in a particular embodiment, in addition to the component polypeptides, the composition of the invention may comprise or contain at least one fat-soluble vitamin, and/or at least one water-soluble vitamin, and/or at least one trace mineral. Also at least one macro mineral may be included.

Examples of fat-soluble vitamins are vitamin A, vitamin D3, vitamin E, and vitamin K, e.g. vitamin K3.

Examples of water-soluble vitamins are vitamin B12, biotin and choline, vitamin B1, vitamin B2, vitamin B6, niacin, folic acid and panthothenate, e.g. Ca-D-panthothenate.

Examples of trace minerals are manganese, zinc, iron, copper, iodine, selenium, and cobalt.

Examples of macro minerals are calcium, phosphorus and sodium.

Further, optional, feed-additive ingredients are colouring agents, e.g. carotenoids such as beta-carotene, astaxanthin, and lutein; aroma compounds; stabilizers; polyunsaturated fatty acids; reactive oxygen generating species; antimicrobial peptides; and/or at least one additional enzyme.

Additional enzyme components of the invention include at least one polypeptide having amylase, preferably alpha-amylase, activity, and/or at least one polypeptide having xylanase activity; and/or at least one polypeptide having endoglucanase activity; and/or at least one polypeptide having endo-1,3(4)-beta-glucanase activity; and/or at least one polypeptide having phytase activity; and/or at least one polypeptide having galactanase activity; and/or at least one polypeptide having alpha-galactosidase activity; and/or at least one other polypeptide having protease activity (EC 3.4.-.-); and/or at least one polypeptide having phospholipase A1 (EC 3.1.1.32), phospholipase A2 (EC 3.1.1.4), lysophospholipase (EC 3.1.1.5), phospholipase C (EC 3.1.4.3), and/or phospholipase D (EC 3.1.4.4) activity.

Alpha-amylase activity can be measured as is known in the art, e.g. using a starch-based substrate.

Xylanase activity can be measured using any assay, in which a substrate is employed, that includes 1,4-beta-D-xylosidic endo-linkages in xylans. Different types of substrates are available for the determination of xylanase activity e.g. Xylazyme cross-linked arabinoxylan tablets (from MegaZyme), or insoluble powder dispersions and solutions of azo-dyed arabinoxylan.

Endoglucanase activity can be determined using any endoglucanase assay known in the art. For example, various cellulose- or beta-glucan-containing substrates can be applied. An endoglucanase assay may use AZCL-Barley beta-Glucan, or preferably (1) AZCL-HE-Cellulose, or (2) Azo-CM-cellulose as a substrate. In both cases, the degradation of the substrate is followed spectrophotometrically at OD_{595} (see the Megazyme method for AZCL-polysaccharides for the assay of endo-hydrolases at http://www.megazyme.com/booklets/AZCLPOL.pdf.

Endo-1,3(4)-beta-glucanase activity can be determined using any endo-1,3(4)-beta-glucanase assay known in the art. A preferred substrate for endo-1,3(4)-beta-glucanase activity measurements is a cross-linked azo-coloured beta-glucan Barley substrate, wherein the measurements are based on spectrophotometric determination principles.

Phytase activity can be measured using any suitable assay, e.g. the FYT assay described in Example 4 of WO 98/28408.

Galactanase can be assayed e.g. with AZCL galactan from Megazyme, and alpha-galactosidase can be assayed e.g. with pNP-alpha-galactoside.

For assaying these enzyme activitites the assay-pH and the assay-temperature are to be adapted to the enzyme in question (preferably a pH close to the optimum pH, and a temperature close to the optimum temperature). A preferred assay pH is in the range of 2-10, preferably 3-9, more preferably pH 3 or 4 or 5 or 6 or 7 or 8, for example pH 3 or pH 7. A preferred assay temperature is in the range of 20-90°C, preferably 30-90°C, more preferably 40-80°C, even more preferably 40-70°C, preferably 40 or 45 or 50°C. The enzyme activity is

defined by reference to appropriate blinds, e.g. a buffer blind.

Examples of antimicrobial peptides (AMP's) are CAP18, Leucocin A, Tritrpticin, Protegrin-1, Thanatin, Defensin, Lactoferrin, Lactoferricin, and Ovispirin such as Novispirin (Robert Lehrer, 2000), Plectasins, and Statins, including the compounds and polypeptides disclosed in WO 03/044049 and WO 03/048148, as well as variants or fragments of the above that retain antimicrobial activity.

Examples of antifungal polypeptides (AFP's) are the Aspergillus giganteus, and Aspergillus niger peptides, as well as variants and fragments thereof which retain antifungal activity, as disclosed in WO 94/01459 and WO 02/090384.

Examples of polyunsaturated fatty acids are C18, C20 and C22 polyunsaturated fatty acids, such as arachidonic acid, docosohexaenoic acid, eicosapentaenoic acid and gamma-linoleic acid.

Examples of reactive oxygen generating species are chemicals such as perborate, persulphate, or percarbonate; and enzymes such as an oxidase, an oxygenase or a syntethase.

Usally fat and water soluble vitamins, as well as trace minerals form part of a so-called premix intended for addition to the feed, whereas macro minerals are usually separately added to the feed. A premix enriched with a protease of the invention, is an example of an animal feed additive of the invention.

In a particular embodiment, the animal feed additive of the invention is intended for being included (or prescribed as having to be included) in animal diets or feed at levels of 0.01 to 10.0%; more particularly 0.05 to 5.0%; or 0.2 to 1.0% (% meaning g additive per 100 g feed). This is so in particular for premixes.

The nutritional requirements of these components (exemplified with poultry and piglets/pigs) are listed in Table A of WO 01/58275. Nutritional requirement means that these components should be provided in the diet in the concentrations indicated.

In the alternative, the animal feed additive of the invention comprises at least one of the individual components specified in Table A of WO 01/58275. At least one means either of, one or more of, one, or two, or three, or four and so forth up to all thirteen, or up to all fifteen individual components. More specifically, this at least one individual component is included in the additive of the invention in such an amount as to provide an in-feed-concentration within the range indicated in column four, or column five, or column six of Table A.

The present invention also relates to animal feed compositions. Animal feed compositions or diets have a relatively high content of protein. Poultry and pig diets can be characterised as indicated in Table B of WO 01/58275, columns 2-3. Fish diets can be characterised as indicated in column 4 of this Table B. Furthermore such fish diets usually have a crude fat content of 200-310 g/kg. WO 01/58275 corresponds to US 09/779334 which

is hereby incorporated by reference.

An animal feed composition according to the invention has a crude protein content of 50-800 g/kg, and furthermore comprises at least one protease variant as claimed herein.

Furthermore, or in the alternative (to the crude protein content indicated above), the animal feed composition of the invention has a content of metabolisable energy of 10-30 MJ/kg; and/or a content of calcium of 0.1-200 g/kg; and/or a content of available phosphorus of 0.1-200 g/kg; and/or a content of methionine of 0.1-100 g/kg; and/or a content of methionine plus cysteine of 0.1-150 g/kg; and/or a content of lysine of 0.5-50 g/kg.

In particular embodiments, the content of metabolisable energy, crude protein, calcium, phosphorus, methionine, methionine plus cysteine, and/or lysine is within any one of ranges 2, 3, 4 or 5 in Table B of WO 01/58275 (R. 2-5).

Crude protein is calculated as nitrogen (N) multiplied by a factor 6.25, i.e. Crude protein (g/kg)= N (g/kg) x 6.25. The nitrogen content is determined by the Kjeldahl method (A.O.A.C., 1984, Official Methods of Analysis 14th ed., Association of Official Analytical Chemists, Washington DC).

Metabolisable energy can be calculated on the basis of the NRC publication Nutrient requirements in swine, ninth revised edition 1988, subcommittee on swine nutrition, committee on animal nutrition, board of agriculture, national research council. National Academy Press, Washington, D.C., pp. 2-6, and the European Table of Energy Values for Poultry Feed-stuffs, Spelderholt centre for poultry research and extension, 7361 DA Beekbergen, The Netherlands. Grafisch bedrijf Ponsen & looijen bv, Wageningen. ISBN 90-71463-12-5.

The dietary content of calcium, available phosphorus and amino acids in complete animal diets is calculated on the basis of feed tables such as Veevoedertabel 1997, gegevens over chemische samenstelling, verteerbaarheid en voederwaarde van voedermiddelen, Central Veevoederbureau, Runderweg 6, 8219 pk Lelystad. ISBN 90-72839-13-7.

In a particular embodiment, the animal feed composition of the invention contains at least one protein. The protein may be an animal protein, such as meat and bone meal, and/or fish meal; or, in a particular embodiment, it may be a vegetable protein. The term vegetable proteins as used herein refers to any compound, composition, preparation or mixture that includes at least one protein derived from or originating from a vegetable, including modified proteins and protein-derivatives. In particular embodiments, the protein content of the vegetable proteins is at least 10, 20, 30, 40, 50, or 60% (w/w).

Vegetable proteins may be derived from vegetable protein sources, such as legumes and cereals, for example materials from plants of the families Fabaceae (Leguminosae), Cruciferaceae, Chenopodiaceae, and Poaceae, such as soy bean meal, lupin meal and rapeseed meal.

In a particular embodiment, the vegetable protein source is material from one or more plants of the family Fabaceae, e.g. soybean, lupine, pea, or bean.

In another particular embodiment, the vegetable protein source is material from one or more plants of the family Chenopodiaceae, e.g. beet, sugar beet, spinach or quinoa.

Other examples of vegetable protein sources are rapeseed, sunflower seed, cotton seed, and cabbage.

Soybean is a preferred vegetable protein source.

Other examples of vegetable protein sources are cereals such as barley, wheat, rye, oat, maize (corn), rice, triticale, and sorghum.

In still further particular embodiments, the animal feed composition of the invention contains 0-80% maize; and/or 0-80% sorghum; and/or 0-70% wheat; and/or 0-70% Barley; and/or 0-30% oats; and/or 0-40% soybean meal; and/or 0-25%, preferably 0-10%, fish meal; 0-25% meat and bone meal; and/or 0-20% whey.

Animal diets can e.g. be manufactured as mash feed (non pelleted) or pelleted feed. Typically, the milled feed-stuffs are mixed and sufficient amounts of essential vitamins and minerals are added according to the specifications for the species in question. Enzymes can be added as solid or liquid enzyme formulations. For example, a solid enzyme formulation is typically added before or during the mixing step; and a liquid enzyme preparation is typically added after the pelleting step. The enzyme may also be incorporated in a feed additive or premix.

The final enzyme concentration in the diet is within the range of 0.01-200 mg enzyme protein per kg diet, for example in the range of 0.5-25 mg enzyme protein per kg animal diet.

The protease variant should of course be applied in an effective amount, i.e. in an amount adequate for improving solubilisation and/or improving nutritional value of feed. It is at present contemplated that the enzyme is administered in one or more of the following amounts (dosage ranges): 0.01-200; 0.01-100; 0.5-100; 1-50; 5-100; 10-100; 0.05-50; or 0.10-10 – all these ranges being in mg protease enzyme protein per kg feed (ppm).

For determining mg enzyme protein per kg feed, the protease is purified from the feed composition, and the specific activity of the purified protease is determined using a relevant assay (see under protease activity, substrates, and assays). The protease activity of the feed composition as such is also determined using the same assay, and on the basis of these two determinations, the dosage in mg enzyme protein per kg feed is calculated.

The same principles apply for determining mg enzyme protein in feed additives. Of course, if a sample is available of the protease used for preparing the feed additive or the feed, the specific activity is determined from this sample (no need to purify the protease from the feed composition or the additive).

Detergent Compositions

The protease variant of the invention may be added to and thus become a component of a detergent composition.

The detergent composition of the invention may for example be formulated as a hand or machine laundry detergent composition including a laundry additive composition suitable for pre-treatment of stained fabrics and a rinse added fabric softener composition, or be formulated as a detergent composition for use in general household hard surface cleaning operations, or be formulated for hand or machine dishwashing operations.

In a specific aspect, the invention provides a detergent additive comprising the protease variant of the invention. The detergent additive as well as the detergent composition may comprise one or more other enzymes such as another protease, such as alkaline proteases from Bacillus, a lipase, a cutinase, an amylase, a carbohydrase, a cellulase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, e.g., a laccase, and/or a peroxidase.

In general the properties of the chosen enzyme(s) should be compatible with the selected detergent, (i.e. pH-optimum, compatibility with other enzymatic and non-enzymatic ingredients, etc.), and the enzyme(s) should be present in effective amounts.

Suitable lipases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful lipases include lipases from Humicola (synonym Thermomyces), e.g. from H. lanuginosa (T. lanuginosus) as described in EP 258068 and EP 305216 or from H. insolens as described in WO 96/13580, a Pseudomonas lipase, e.g. from P. alcaligenes or P. pseudoalcaligenes (EP 218272), P. cepacia (EP 331376), P. stutzeri (GB 1,372,034), P. fluorescens, Pseudomonas sp. strain SD 705 (WO 95/06720 and WO 96/27002), P. wisconsinensis (WO 96/12012), a Bacillus lipase, e.g. from B. subtilis (Dartois et al. (1993), Biochemica et Biophysica Acta, 1131, 253-360), B. stearothermophilus (JP 64/744992) or B. pumilus (WO 91/16422). Other examples are lipase variants such as those described in WO 92/05249, WO 94/01541, EP 407225, EP 260105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 and WO 97/07202. Preferred commercially available lipase enzymes include LipolaseTM and Lipolase UltraTM (Novozymes A/S).

Suitable amylases (alpha- and/or beta-) include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Amylases include, for example, alpha-amylases obtained from Bacillus, e.g. a special strain of B. licheniformis, described in more detail in GB 1,296,839. Examples of useful amylases are the variants described in WO 94/02597, WO 94/18314, WO 95/26397, WO 96/23873, WO 97/43424, WO 00/60060, and WO 01/66712, especially the variants with substitutions in one or more of the following positions: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208,

209, 243, 264, 304, 305, 391, 408, and 444. Commercially available amylases are Natalase™, Supramyl™, Stainzyme™, Duramyl™, Termamyl™, Fungamyl™ and BAN™ (Novozymes A/S), Rapidase™ and Purastar™ (from Genencor International Inc.).

Suitable cellulases include those of bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Suitable cellulases include cellulases from the genera Bacillus, Pseudomonas, Humicola, Fusarium, Thielavia, Acremonium, e.g. the fungal cellulases produced from Humicola insolens, Myceliophthora thermophila and Fusarium oxysporum disclosed in US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 and WO 89/09259. Especially suitable cellulases are the alkaline or neutral cellulases having colour care benefits. Examples of such cellulases are cellulases described in EP 0 495257, EP 531372, WO 96/11262, WO 96/29397, WO 98/08940. Other examples are cellulase variants such as those described in WO 94/07998, EP 0 531 315, US 5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 and WO 99/01544. Commercially available cellulases include CelluzymeTM, and CarezymeTM (Novozymes A/S), ClazinaseTM, and Puradax HATM (Genencor International Inc.), and KAC-500(B)TM (Kao Corporation).

Suitable peroxidases/oxidases include those of plant, bacterial or fungal origin. Chemically modified or protein engineered mutants are included. Examples of useful peroxidases include peroxidases from Coprinus, e.g. from C. cinereus, and variants thereof as those described in WO 93/24618, WO 95/10602, and WO 98/15257. Commercially available peroxidases include GuardzymeTM (Novozymes).

The detergent enzyme(s) may be included in a detergent composition by adding separate additives containing one or more enzymes, or by adding a combined additive comprising all of these enzymes. A detergent additive of the invention, i.e. a separate additive or a combined additive, can be formulated e.g. as a granulate, a liquid, a slurry, etc. Preferred detergent additive formulations are granulates, in particular non-dusting granulates, liquids, in particular stabilized liquids, or slurries.

Non-dusting granulates may be produced, e.g., as disclosed in US 4,106,991 and 4,661,452 and may optionally be coated by methods known in the art. Examples of waxy coating materials are poly(ethylene oxide) products (polyethyleneglycol, PEG) with mean molar weights of 1000 to 20000; ethoxylated nonylphenols having from 16 to 50 ethylene oxide units; ethoxylated fatty alcohols in which the alcohol contains from 12 to 20 carbon atoms and in which there are 15 to 80 ethylene oxide units; fatty alcohols; fatty acids; and mono- and di- and triglycerides of fatty acids. Examples of film-forming coating materials suitable for application by fluid bed techniques are given in GB 1483591. Liquid enzyme preparations may, for instance, be stabilized by adding a polyol such as propylene glycol, a sugar or sugar alcohol, lactic acid or boric acid according to established methods. Protected enzymes may be prepared according to the method disclosed in EP 238216.

The detergent composition of the invention may be in any convenient form, e.g., a bar, a tablet, a powder, a granule, a paste or a liquid. A liquid detergent may be aqueous, typically containing up to 70 % water and 0-30 % organic solvent, or non-aqueous.

The detergent composition comprises one or more surfactants, which may be non-ionic including semi-polar and/or anionic and/or cationic and/or zwitterionic. The surfactants are typically present at a level of from 0.1% to 60% by weight.

When included therein the detergent will usually contain from about 1% to about 40% of an anionic surfactant such as linear alkylbenzenesulfonate, alpha-olefinsulfonate, alkyl sulfate (fatty alcohol sulfate), alcohol ethoxysulfate, secondary alkanesulfonate, alpha-sulfo fatty acid methyl ester, alkyl- or alkenylsuccinic acid or soap.

When included therein the detergent will usually contain from about 0.2% to about 40% of a non-ionic surfactant such as alcohol ethoxylate, nonylphenol ethoxylate, alkylpolyglycoside, alkyldimethylamineoxide, ethoxylated fatty acid monoethanolamide, fatty acid monoethanolamide, polyhydroxy alkyl fatty acid amide, or N-acyl N-alkyl derivatives of glucosamine ("glucamides").

The detergent may contain 0-65 % of a detergent builder or complexing agent such as zeolite, diphosphate, triphosphate, phosphonate, carbonate, citrate, nitrilotriacetic acid, ethylenediaminetetraacetic acid, diethylenetriaminepentaacetic acid, alkyl- or alkenylsuccinic acid, soluble silicates or layered silicates (e.g. SKS-6 from Hoechst).

The detergent may comprise one or more polymers. Examples are carboxymethylcellulose, poly(vinylpyrrolidone), poly (ethylene glycol), poly(vinyl alcohol), poly(vinylpyridine-N-oxide), poly(vinylimidazole), polycarboxylates such as polyacrylates, maleic/acrylic acid copolymers and lauryl methacrylate/acrylic acid copolymers.

The detergent may contain a bleaching system which may comprise a H2O2 source such as perborate or percarbonate which may be combined with a peracid-forming bleach activator such as tetraacetylethylenediamine or nonanoyloxybenzenesulfonate. Alternatively, the bleaching system may comprise peroxyacids of e.g. the amide, imide, or sulfone type.

The enzyme(s) of the detergent composition of the invention may be stabilized using conventional stabilizing agents, e.g., a polyol such as propylene glycol or glycerol, a sugar or sugar alcohol, lactic acid, boric acid, or a boric acid derivative, e.g., an aromatic borate ester, or a phenyl boronic acid derivative such as 4-formylphenyl boronic acid, and the composition may be formulated as described in e.g. WO 92/19709 and WO 92/19708.

The detergent may also contain other conventional detergent ingredients such as e.g. fabric conditioners including clays, foam boosters, suds suppressors, anti-corrosion agents, soil-suspending agents, anti-soil redeposition agents, dyes, bactericides, optical brighteners, hydrotropes, tarnish inhibitors, or perfumes.

It is at present contemplated that in the detergent compositions any enzyme, in

particular the enzyme of the invention, may be added in an amount corresponding to 0.01-100 mg of enzyme protein per liter of wash liqour, preferably 0.05-5 mg of enzyme protein per liter of wash liqour, in particular 0.1-1 mg of enzyme protein per liter of wash liqour.

The enzyme of the invention may additionally be incorporated in the detergent formulations disclosed in WO 97/07202.

Method for Generating Protease Variants

The invention also relates to a method for generating a protease variant of an improved property, the method comprising the following steps:

(a) selecting a parent protease of at least 60% identity to amino acids 1 to 188 of SEQ ID NO: 2;

(b) establishing a 3D structure of the parent protease by homology modelling using the Fig. 2 structure as a model; and/or aligning the parent protease according to the alignment of Fig. 1;

(c) proposing at least one amino acid substitution, e.g. by:

(i) subjecting the 3D structure of (b) to MD simulations at increased temperatures, and identifying regions in the amino acid sequence of the parent protease of high mobility (isotropic fluctuations);

(ii) introducing disulfid bridges by way of cysteine substitutions (C-C);

(iii) introducing proline substitutions (P);

(iv) replacing exposed neutral amino acid residues with negatively charged amino acid residues (E,D);

(v) replacing exposed neutral amino aicd residues with positively charged amino acid residues (R,K);

(vi) replacing small amino acid residues inside the protein with bulkier amino acid residues (W);

(vii) comparing by homology alignment and/or homology modelling according to step (c)(i) at least two related parent proteases and transferring amino acid residue differences inbetween these protease backbones, preferably from a backbone having the improved property to a backbone not having this improved property;

(d) preparing a DNA sequence encoding the parent protease but for inclusion of a DNA codon of the at least one amino acid substitution proposed in steps (c)(ii)-(c)(vii), or subjecting the parent DNA sequence to random mutagenesis, targetting at least one of the regions identified in step (c)(i);

(e) expressing the DNA sequence obtained in step (d) in a host cell, and

(h) selecting a host cell expressing a protease variant with an improved property.

The invention furthermore relates to a method for producing a protease variant obtainable or obtained by the method of generating protease variants described above, comprising (a) cultivating the host cell to produce a supernatant comprising the variant; and (b) recovering the variant.

The invention also relates to isolated nucleic acid sequences comprising a nucleic acid sequence which encodes the protease variant obtainable according to this method, as well as methods for producing it by (a) cultivating the host cell to produce a supernatant comprising the variant; and (b) recovering the variant; a transgenic plant, or plant part, capable of expressing it; transgenic, non-human animals, or products, or elements thereof, being capable of expressing it; animal feeds, as well as animal feed additives, comprising it; methods for improving the nutritional value of an animal feed by use thereof; methods for the treatment of proteins, such as vegetable proteins, by use thereof; as well s the use thereof (i) in animal feed; (ii) in the preparation of animal feed; (iii) for improving the nutritional value of animal feed; and/or (iv) for the treatment of proteins; and/or in detergents.

Alternative Embodiment

In an alternative embodiment, the term "alteration" is used instead of "substitution" as the general term for amendments in the protease molecule. This alternative embodiment includes each of the claims formulated as examplified below for claim 1, and also specifically includes everything what is stated herein, e.g. definitions (other than the definition of substitution), i.e. the various aspects, particular embodiments etc.

A variant of a parent protease, comprising an alteration in at least one position of at least one region selected from the group of regions consisting of:
6-18; 22-28; 32-39; 42-58; 62-63; 66-76; 78-100; 103-106; 111-114; 118-131; 134-136; 139-141; 144-151; 155-156; 160-176; 179-181; and 184-188; wherein

(a) the alteration(s) are independently

(i) an insertion of an amino acid immediately downstream of the position,

(ii) a deletion of the amino acid which occupies the position, and/or

(iii) a substitution of the amino acid which occupies the position;

(b) the variant has protease activity; and

(c) each position corresponds to a position of SEQ ID NO: 2, preferably amino acids 1 to 188 thereof; and

(d) the variant has a percentage of identity to SEQ ID NO: 2, preferably to amino acids 1 to 188 thereof, of at least 60%.

The term "polypeptide variant", "protein variant", "enzyme variant", "protease variant" or simply "variant" refers to a polypeptide of the invention comprising one or more alteration(s),

such as substitution(s), insertion(s), deletion(s), and/or truncation(s) of one or more specific amino acid residue(s) in one or more specific position(s) in the polypeptide.

The term "parent polypeptide", "parent protein", "parent enzyme", "standard enzyme", "parent protease" or simply "parent" refers to the polypeptide on which the variant was based. This term also refers to the polypeptide with which a variant is compared and aligned.

The term "randomized library", "variant library", or simply "library" refers to a library of variant polypeptides. Diversity in the variant library can be generated via mutagenesis of the genes encoding the variants at the DNA triplet level, such that individual codons are variegated e.g. by using primers of partially randomized sequence in a PCR reaction. Several techniques have been described, by which one can create a diverse combinatorial library by variegating several nucleotide positions in a gene and recombining them, for instance where these positions are too far apart to be covered by a single (spiked or doped) oligonucleotide primer. These techniques include the use of in vivo recombination of the individually diversified gene segments as described in WO 97/07205 on page 3, lines 8 to 29 (Novozymes A/S). They also include the use of DNA shuffling techniques to create a library of full length genes, wherein several gene segments are combined, and wherein each segment may be diversified e.g. by spiked mutagenesis (Stemmer, Nature 370, pp. 389-391, 1994 and US 5,811,238; US 5,605,793; and US 5,830,721). One can use a gene encoding a protein "backbone" (wildtype parent polypeptide) as a template polynucleotide, and combine this with one or more single or double-stranded oligonucleotides as described in WO 98/41623 and in WO 98/41622 (Novozymes A/S). The single-stranded oligonucleotides could be partially randomized during synthesis. The double-stranded oligonucleotides could be PCR products incorporating diversity in a specific region. In both cases, one can dilute the diversity with corresponding segments encoding the sequence of the backbone protein in order to limit the average number of changes that are introduced.

Methods have also been established for designing the ratios of nucleotide mixtures (A; C; T; G) to be inserted in specific codon positions during oligo- or polynucleotide synthesis, so as to introduce a bias in order to approximate a desired frequency distribution towards a set of one or more desired amino acids that will be encoded by the particular codons. It may be of interest to produce a variant library, that comprises permutations of a number of known amino acid modifications in different locations in the primary sequence of the polypeptide. These could be introduced post-translationally or by chemical modification sites, or they could be introduced through mutations in the encoding genes. The modifications by themselves may previously have been proven beneficial for one reason or another (e.g. decreasing antigenicity, or improving specific activity, performance, stability, or other characteristics). In such instances, it may be desirable first to create a library of diverse combinations of known sequences. For example, if twelwe individual mutations are known, one could combine (at

least) twelwe segments of the parent protein encoding gene, wherein each segment is present in two forms: one with, and one without the desired mutation. By varying the relative amounts of those segments, one could design a library (of size 212) for which the average number of mutations per gene can be predicted. This can be a useful way of combining mutations, that by themselves give some, but not sufficient effect, without resorting to very large libraries, as is often the case when using 'spiked mutagenesis'. Another way to combine these 'known mutations' could be by using family shuffling of oligomeric DNA encoding the known mutations with fragments of the full length wild type sequence.

In describing the various variants produced or contemplated according to the invention, a number of nomenclatures and conventions are used which are described in detail below. A frame of reference is first defined by aligning the variant polypeptide with a parent enzyme. A preferred parent enzyme is Protease 10 (amino acids 1 to 188 of SEQ ID NO: 2). Thereby a number of alterations will be defined in relation to the amino acid sequence of amino acids 1 to 188 of SEQ ID NO: 2.

A substitution in a variant is indicated as:

Original amino acid - position - substituted amino acid;

The three or one letter codes are used, including the codes Xaa and X to indicate any amino acid residue. Accordingly, the notation "T82S" or "Thr82Ser" means, that the variant comprises a substitution of threonine with serine in the variant amino acid position corresponding to the amino acid in position 82 in the parent enzyme, when the two are aligned as indicated above.

Where the original amino acid residue may be any amino acid residue, a short hand notation may at times be used indicating only the position, and the substituted amino acid, for example:

Position - substituted amino acid; or "82S",

Such a notation is particular relevant in connection with modification(s) in a series of homologous polypeptides.

Similarly when the identity of the substituting amino acid residue(s) is immaterial:

Original amino acid - position; or "T82"

When both the original amino acid(s) and substituted amino acid(s) may be any amino acid, then only the position is indicated, e.g.: "82".

When the original amino acid(s) and/or substituted amino acid(s) may comprise more than one, but not all amino acid(s), then the amino acids are listed separated by commas:

Original amino acids - position no. - substituted amino acids; or "T10E,D,Y".

A number of examples of this nomenclature are listed below:

The substitution of threonine for histidine in position 91 is designated as: "His91Thr" or "H91T"; or the substitution of any amino acid residue acid for histidine in position 91 is designated as: "His91Xaa" or "H91X" or "His91" or "H91".

For a modification where the original amino acid(s) and/or substituted amino acid(s) may comprise more than one, but not all amino acid(s), the substitution of glutamic acid, aspartic acid, or tyrosine for threonine in position 10:

"Thr10Glu,Asp,Tyr" or "T10E,D,Y"; which indicates the specific variants: "T10E", "T10D", and "T10Y".

A deletion of glycine in position 26 will be indicated by: "Gly26*" or "G26*"

Correspondingly, the deletion of more than one amino acid residue, such as the deletion of glycine and glutamine in positions 26 and 27 will be designated "Gly26*+Gln27*" or "G26*+Q27*"

The insertion of an additional amino acid residue such as e.g. a lysine after G26 is indicated by: "Gly26GlyLys" or "G26GK"; or, when more than one amino acid residue is inserted, such as e.g. a Lys, and Ala after G26 this will be indicated as: "Gly26GlyLysAla" or "G26GKA".

In such cases the inserted amino acid residue(s) are numbered by the addition of lower case letters to the position number of the amino acid residue preceding the inserted amino acid residue(s). In the above example the sequences would thus be:

Parent:	Variant:		
26	26	26a	26b
G	G	K	A

In cases where an amino acid residue identical to the existing amino acid residue is inserted, it is clear that degeneracy in the nomenclature arises. If for example a glycine is inserted after the glycine in the above example this would be indicated by "G26GG".

Given that an alanine were present in position 25, the same actual change could just as well be indicated as "A25AG":

	Parent:		Variant:		
Numbering I:	25	26	25	26	26a
Sequence:	A	G	A	G	G
Numbering II:			25	25a	26

Such instances will be apparent to the skilled person, and the indication "G26GG" and corresponding indications for this type of insertions is thus meant to comprise such equivalent degenerate indications.

By analogy, if amino acid sequence segments are repeated in the parent polypeptide and/or in the variant, it will be apparent to the skilled person that equivalent degenerate indications are comprised, also when other alterations than insertions are listed such as

deletions and/or substitutions. For instance, the deletion of two consecutive amino acids "AG" in the sequence "AGAG" from position 194-197, may be written as "A194*+G1956*" or "A196*+G197*":

	Parent:				Variant:	
Numbering I:	194	195	196	197	194	195
Sequence:	A	G	A	G	A	G
Numbering II:					196	197

Variants comprising multiple modifications are separated by pluses, e.g.: "Arg170Tyr+Gly195Glu" or "R170Y+G195E", representing modifications in positions 170 and 195 substituting tyrosine and glutamic acid for arginine and glycine, respectively. Thus, "Tyr167Gly,Ala,Ser,Thr+Arg170Gly,Ala,Ser,Thr" designates the following variants: "Tyr167Gly+Arg170Gly", "Tyr167Gly+Arg170Ala", "Tyr167Gly+Arg170Ser", "Tyr167Gly+Arg170Thr", "Tyr167Ala+Arg170Gly", "Tyr167Ala+Arg170Ala", "Tyr167Ala+Arg170Ser", "Tyr167Ala+Arg170Thr", "Tyr167Ser+Arg170Gly", "Tyr167Ser+Arg170Ala", "Tyr167Ser+Arg170Ser", "Tyr167Ser+Arg170Thr", "Tyr167Thr+Arg170Gly", "Tyr167Thr+Arg170Ala", "Tyr167Thr+Arg170Ser", and "Tyr167Thr+Arg170Thr".

This nomenclature is particular relevant relating to modifications aimed at substituting, inserting or deleting amino acid residues having specific common properties, such modifications are referred to as conservative amino acid modification(s).

Various embodiments

These are additional various embodiments of the invention:

The variant of any one of claims 1-16 and 18-20 which comprises at least one of the following substitutions: T10Y, A24S, V51T, E53Q, T82S, A86Q, T87S, I96A, G118N, S122R, N130S, L186I.

The variant of any one of claims 1-16 and 18-19 which comprises at least one of the following substitutions: R38T; Q42G,P; R49T,Q; Q54N,R; A89S,T; H91S,T; N92S; S99A,Q; A120T; E125Q; T129Y,Q; M131L; T135N; Y147F; N151S; R165S; T166V,F; F171Y; V179I,L; preferably at least one of the following substitutions: R38T; N92S; A120T; E125Q; M131L; T135N; Y147F; N151S; R165S; and/or F171Y.

The variant of any one of claims 1-19 which comprises at least one of the following substitutions: A25S, T44S, A62S, P95A, V100I, I114V, T176N, N180S, V184L, R185T.

The variant of any one of claims 1-20 which has amended properties, such as an improved thermostability and/or a higher or lower optimum temperature, such as a Tm of at least 83.1°C as measured by DSC in 10mM sodium phosphate, 50 mM sodium chloride, pH 7.0.

The variant of any one of claims 1-20 which derives from a strain of the genus Nocardiopsis, such as Nocardiopsis alba, Nocardiopsis antarctica, Nocardiopsis prasina, Nocardiopsis composta, Nocardiopsis dassonvillei, Nocardiopsis exhalans, Nocardiopsis halophila, Nocardiopsis halotolerans, Nocardiopsis kunsanensis, Nocardiopsis listeri, Nocardiopsis lucentensis, Nocardiopsis metallicus, Nocardiopsis sp., Nocardiopsis synnemataformans, Nocardiopsis trehalosi, Nocardiopsis tropica, Nocardiopsis umidischolae, or Nocardiopsis xinjiangensis, preferably Nocardiopsis alba DSM 15647, Nocardiopsis dassonvillei NRRL 18133, Nocardiopsis dassonvillei subsp. dassonvillei DSM 43235, Nocardiopsis prasina DSM 15648, Nocardiopsis prasina DSM 15649, Nocardiopsis sp. NRRL 18262, most preferably Nocardiopsis sp. FERM P-18676.

A composition, such as an animal feed additive, comprising at least one protease variant of any one of claims 1-20, and

(a) at least one fat soluble vitamin;

(b) at least one water soluble vitamin; and/or

(c) at least one trace mineral,

optionally further comprising at least one enzyme selected from the following group of enzymes: amylases, galactanases, alpha-galactosidases, xylanases, endoglucanases, endo-1,3(4)-beta-glucanases, phytases, phospholipases, and other proteases; if desired also comprising at least one amylase, and/or phospholipase.

The present invention is further described by the following examples which should not be construed as limiting the scope of the invention.

Examples

Example 1: Protease assays

<u>pNA assay</u>

pNA substrate : Suc-AAPF-pNA (Bachem L-1400).

Temperature : Room temperature (25°C)

Assay buffers :100mM succinic acid, 100mM HEPES, 100mM CHES, 100mM CABS, 1mM $CaCl_2$, 150mM KCl, 0.01% Triton X-100 adjusted to pH-values 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, and 12.0 with HCl or NaOH.

20µl protease (diluted in 0.01% Triton X-100) is mixed with 100µl assay buffer. The assay is started by adding 100µl pNA substrate (50mg dissolved in 1.0ml DMSO and further diluted 45x with 0.01% Triton X-100). The increase in OD_{405} is monitored as a measure of the protease activity.

<u>Protazyme AK assay</u>

Substrate : Protazyme AK tablet (cross-linked and dyed casein; from Megazyme)

Temperature : controlled (assay temperature).

Assay buffers :100mM succinic acid, 100mM HEPES, 100mM CHES, 100mM CABS, 1mM $CaCl_2$, 150mM KCl, 0.01% Triton X-100 adjusted to pH-values 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0 and 11.0 with HCl or NaOH.

A Protazyme AK tablet is suspended in 2.0ml 0.01% Triton X-100 by gentle stirring. 500µl of this suspension and 500µl assay buffer are mixed in an Eppendorf tube and placed on ice. 20µl protease sample (diluted in 0.01% Triton X-100) is added. The assay is initiated by transferring the Eppendorf tube to an Eppendorf thermomixer, which is set to the assay temperature. The tube is incubated for 15 minutes on the Eppendorf thermomixer at its highest shaking rate (1400 rpm). The incubation is stopped by transferring the tube back to the ice bath. Then the tube is centrifuged in an icecold centrifuge for a few minutes and 200µl supernatant is transferred to a microtiter plate. OD_{650} is read as a measure of protease activity. A buffer blind is included in the assay (instead of enzyme).

Example 2: Preparation and Testing of Protease Variants

Four protease variants comprising the amino acid sequence of amino acids 1 to 188 of SEQ ID NO: 2 (Protease 10) with the single substitutions N47D, T127R, N92K, and Q54R, respectively, were prepared as described below for variant N47D.

Site directed mutagenesis was carried out using the Mega-primer method as described by Sarkar and Sommer, 1990 (BioTechniques 8: 404-407).

The N47D variant was constructed by use of the following primers, of which primer R10WT-CL29 (SEQ ID NO: 11) is gene specific, and primer RSWT126 (SEQ ID NO: 12) mutagenic:

R10WT-CL29: 5' CCGATTATGGAGCGGATTGAACATGCG 3' (SEQ ID NO: 11)

RSWT126: 5' GTGACCATCGGCGACGGCAGGGGCGTCTTCG 3' (SEQ ID NO: 12),

to amplify by PCR an approximately 469 bp DNA fragment from the construct described below.

The Protease 10 DNA construct used for the above amplification was an expression cassette (SEQ ID NO: 13) for incorporation into the genome of Bacillus subtilis. The construct contains a fusion of DNA encoding the signal sequence and the gene encoding the pro- and the mature protein of Protease 10 (SEQ ID NO: 14), a promoter construction, and also the cat gene conferring resistance towards chloramphenicol. To facilitate the integration into the genome by homologous recombination, flanking regions of around 3 kb of a Bacillus subtilis endogenous genes were incorporated upstream and downstream of the Protease 10 encoding sequence.

The resulting 469 bp fragment was purified from an agarose gel (Sigma Aldrich cat.no. A6877) and used as a Mega-primer together with primer R10WT-CL39N (SEQ ID NO: 15) in a

second PCR carried out on the same template.

R10WT-CL39N: 5' GGAGCTCTGAAAAAAGGAGAGGATAAAGAATGAA 3' (SEQ ID NO: 15).

The full construction of approximately 10kb is made in vitro by long range PCR, using the oligonucleotides R10WT-CL28N (SEQ ID NO: 16), R10WT-CL28C (SEQ ID NO: 17), and the Expand Long Template PCR System from Roche Applied Science (cat no. 11759060), according to the suppliers manual.

R10WT-CL28N: 5' GCGTTCCGATAATCGCGGTGACAATGCCG 3' (SEQ ID NO: 16)

R10WT-CL28C: 5' TTCATGAGTCTGCGCCCTGAGATCCTCTG 3' (SEQ ID NO: 17)

The resulting approximately 1.2 kb fragment was purified and combined in a new PCR reaction using Expand Long Template PCR System with the flanking fragments of the construction made by two PCR reactions using R10WT-2C-rev (SEQ ID NO: 18) and R10WT-CL28C (SEQ ID NO: 17); and RSWT001 (SEQ ID NO: 19) and R10WT-CL28N (SEQ ID NO: 16) as primer sets. The resulting 10kb fragment can be amplified using the R10WT-CL28N (SEQ ID NO: 16) and R10WT-CL28C (SEQ ID NO: 17) primers, to increase the number of transformants.

R10WT-2C-rev: 5' TAATCGCATGTTCAATCCGCTCCATAATCG 3' (SEQ ID NO: 18)

RSWT001: 5' CCCAACGGTTTCTTCATTCTTTATCCTCTCCTTTTTTTCAGAGC 3' (SEQ ID NO: 19)

Competent cells of an amylase- and protease-low strain of Bacillus subtilis (such as strain SHA273 described in WO92/11357 and WO95/10603) were transformed with the respective resulting PCR fragments, and chlorampenicol resistant transformants were selected and checked by DNA sequencing to verify the presence of the correct mutation on the genome.

Cells of Bacillus subtilis harbouring constructs encoding Protease 10 and each of the four variants thereof were used to incubate shakeflasks containing a rich media (PS-1: 100 g/L Sucrose (Danisco cat.no. 109-0429), 40 g/L crust soy, 10g/L $Na_2HPO_4.12H_2O$ (Merck cat.no. 6579), 0.1ml/L Pluronic PE 6100 (BASF 102-3098)), and cultivation took place for five days at 30°C under vigorous shaking.

After cultivation, the supernatants were diluted four times in a 0.2M Na_2HPO_4 buffer, titrated with a 0.1M citric acid to either pH 4.0 or pH 6.0, and split in two. One half was incubated for four hours at 65°C at the respective pH, after which it was frozen. The other half was frozen immediately and served as the control.

Prior to measuring the residual protease activity, the samples were diluted ten times in 50mM CHES-HEPES buffer, pH 8.5. The activity was determined using a modified version of the Protazyme AK assay of Example 1, solubilising one tablet of the substrate in 4 ml CHES-HEPES buffer, pH 8.5, mixing under continuous agitation one ml of this substrate solution with

20ul of diluted protease sample, which was then incubated at 37°C. The substrate should have the correct temperature prior to adding protease. After 15 minutes the reaction was stopped by adding 100ul 1M NaOH and the insoluble substrate was precipitated by centrifugation at 15000 rpm for 3 minutes after which the absorbance at 650nm was measured. The values should be below OD 3.0, alternatively the protease sample should be diluted more than ten times prior to the activity measurement.

The relative residual activity (%) is calculated by dividing the activity after incubation at 65°C with the activity of the corresponding control. The results of Table 1 below show that all four variants are of an improved thermostability as compared to Protease 10.

Table 1
Residual activity after incubation for four hours at 65°C

Protease	% Residual Actitivty pH 6	% Residual Activity pH 4
Protease 10 + N47D	44	68
Protease 10 + T127R	-	77
Protease 10 + N92K	-	55
Protease 10 + Q54R	52	67
Protease 10	19	41

Example 3: Protease variant 22

A protease variant designated "Protease 22" and comprising a number of substitutions in thirteen of the seventeen regions specified in claim 1 was designed. This variant comprises the following substitutions as compared to the mature part of Protease 10 (amino acids 1-188 of SEQ ID NO: 2): T10Y, A25S, R38T, Q42P, T44S, R49K, Q54R, V56I, A62S, T82S, S99A, G118Ns, S120T, S122R, E125Q, T129Y, N130S, M131L, R165S, T166A, F171Y, T176N, V179L, N180S, V184L, and R185T.

The mature part of Protease 22 is amino acids 1-196 of SEQ ID NO: 21. The DNA sequence corresponding to SEQ ID NO: 21 is SEQ ID NO: 20.

The DNA sequence of SEQ ID NO: 20 was constructed and introduced into a Bacillus host for expression. The expressed protease was purified and characterized as an alpha-lytic protease (peptidase family S1E and/or S2A).

The temperature-activity relationship of Protease 22 was measured at pH9, using the Protazyme AK assay of Example 1, Protease 10 being included for comparative purposes. The results are shown in Table 2 below.

Table 2

Temperature profile at pH9 of Protease 22 and Protease 10

	Relative activity at pH 9	
Temperature (°C)	**Protease 22**	**Protease 10**
15	0.016	0.015
25	0.010	0.024
37	0.028	0.068
50	0.069	0.199
60	0.138	0.510
70	0.474	1.000
80	1.000	0,394
90	0,375	-

From these results it appears that Protease 22 has a higher temperature optimum at pH 9 than the Protease 10, viz. around 80°C as compared to around 70°C.

Differential Scanning Calorimetry (DSC) was used to determine temperature stability at pH 7.0 of Protease 22 and Protease 10. The purified proteases were dialysed over night at 4°C against 10 mM sodium phosphate, 50 mM sodium chloride, pH 7.0 and run on a VP-DSC instrument (Micro Cal) with a constant scan rate of 1.5°C/min from 20 to 100°C. Data-handling was performed using the MicroCal Origin software.

The resulting denaturation or melting temperatures, Tm's, were: For Protease 22: 83.5°C; for Protease 10: 76.5°C.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.